智能情报分析中算法风险及其规制研究*

■ 张涛[1] 马海群[2]

[1] 黑龙江大学信息管理学院 哈尔滨 150080 [2]黑龙江大学信息资源管理研究中心 哈尔滨 150080

摘 要:[目的/意义]近年来,人工智能给国家情报工作带来思维理念和方法技术的转变,使得智能情报分析逐渐成为服务国家情报事业创新发展的重要工作之一。对智能情报分析中的算法风险及其规制进行研究,可以避免人工智能算法所带来的安全风险,减少限制情报工作发展风险方面的因素。[方法/过程]基于对智能情报分析核心算法及研究问题的阐释,结合实际应用场景对算法风险形成原因、产生后果及算法风险各因素间的交互关系进行分析。通过事前评估、事中监管、事后问责的手段建立循序渐进的智能情报分析算法规制框架。[结果/结论]分析"情报"茧房的形成机理,针对智能情报分析中所存在的算法黑箱、算法缺陷、算法操控等风险,提出事前、事中、事后的算法规制良性循环与协调发展建议,并认为正视算法的两面性也是防范与化解算法风险的有效途径。

关键词:智能情报分析 算法风险 算法规制 智能算法 "情报"茧房

分类号:G250 TP18

DOI:10.13266/j.issn.0252-3116.2021.12.004

随着人工智能、大数据、区块链等新兴技术的快速发展,服务于国家安全和发展的情报工作被赋予了新的历史使命,与此同时情报工作在国家治理决策中的功能和作用也发生了变化,在此过程中情报工作的总体目标是为用户制定或提供相对完备的科学性决策[1]。在2019年政府工作报告中提出"智能+"概念后,我国人工智能开启了规模化落地之路,"智能+"情报分析由此产生,它是指利用人工智能技术,依托情报大数据,结合情报工作的规范和方法,合理地为用户提供客观精准情报分析的过程。人工智能时代情报工作人员在复杂多变的决策环境中对海量、异构、多模的数据进行分析时,算法发挥了重要作用,虽然它可以辅助用户完成智慧化的分析过程,提升情报分析效率。同时算法也是一把双刃剑,会引发算法黑箱、算法缺陷、算法操控等一系列安全风险,这也逐渐成为限制情报工作发展的主要因素之一。算法风险在传统情报分析中显现并不突出,它是人工智能视域下情报分析所特有的,而我国在人工智能算法制度建设方面相对薄弱,尤其是在情报工作领域,如不及时防范与化解算法风险不仅会导致情报分析失准,甚至还会给社会稳定乃至国家安全产生造成深远影响。

1 相关研究

智能情报理念源于1993年钱学森先生提出的人机结合是智慧式情报的关键[2]。2015年王飞跃基于钱学森先生的智能情报理念提出情报5.0概念,即平行智能情报[3],此后随着大数据成为新一代的人工智能的重要支撑,将人工智能技术运用于情报分析受到众多学者的关注,现将当前智能情报分析相关成果综述如下。

1.1 相关理论研究

理论研究是智能情报分析的基础。近年来,学界在人工智能与情报工作相结合方面形成了一系列理论层面的研究成果,如:①计算情报研究:计算情报是通过人工智能、脑科学、认知技术来计算情报、辅助决策。李广建等[4-5]提出了计算型情报分析的理论、方法和技术、系统、应用实践;陈雪飞等[6]通过"证据链模型"探索计算情报的实现途径;②数据智能情报研究:胡昌平等[7]在数据智能环境下对情报学理论研究前沿进行分析;栗琳等[8]基于数据智能技术对情报全流程变革及发展进行研究;邱韵霏等[9]分析了数据驱动和知识驱动方法在智能情报分析中的应用与融合发展问题;

* 本文系国家社会科学基金重点项目"总体国家安全观下的国家情报工作制度创新研究"(项目编号:20ATQ004)研究成果之一。

作者简介:张涛(ORCID:0000-0002-3367-4541)副教授,博士研究生,硕士生导师,E-mail:zhangtao@hlju.edu.cn;马海群(ORCID:0000-0002-2091-7620)教授,博士,博士生导师。

收稿日期:2020-11-22 修回日期:2021-02-25 本文起止页码:47-56 本文责任编辑:杜杏叶

③智能情报分析系统：化柏林等[10]对智能情报分析系统的架构设计与关键技术进行研究；曾文等[11]从数据工程视角对提出并构建智能情报分析决策模型；④智能与情报融合研究：孙建军等[12]从思维、资源、技术、教育、路径等方面论述了情报工作发展的“智慧”元素；冯秋燕等[13]构建了新型情报工作体系，从人工智能与情报工作的核心层次上阐述了两者的融合发展；牛海波等[14]对智能时代未来情报工作进行展望。

1.2　相关应用研究

应用研究是智能情报分析的目标。近年来，很多学者将人工智能技术与不同领域情报工作相结合形成了一系列应用研究成果，如：①反恐情报：曾庆华等[15]构建了人机结合、以人为主的智能反恐情报分析系统。②金融情报：丁晓蔚等[16]基于区块链的可信大数据和可信人工智能的理念，提出了新型金融情报分析的理念；③军事情报：王天尧等[17]对人工智能在军事情报工作中的应用现状、特点及启示进行分析；④安全情报：黄云芳等[18]提出智能安全情报分析模型的构建思路；⑤竞争情报：唐晓波等[19]基于人工智能技术对企业竞争情报系统模型进行构建。⑥应急情报：曾子明等[20]构建了突发事件智能情报服务体系，探讨了智能情报服务体系在上海外滩踩踏事件中的具体应用。

虽然学界对智能情报分析展开了一系列研究，但缺乏对智能情报分析中安全风险的研究，尤其是针对算法带来的风险。而在法学、公共管理等其他社科领域，国内外很多学者已经针对算法风险问题展开讨论，在国内，贾开[21]较早提出智能算法在提高人类社会运行效率的同时，也带来了不可解释隐忧、自我强化困境与主体性难题的风险与挑战；张爱军等[22]对人工智能时代算法权力的逻辑、风险及规制进行综合分析；徐凤[23]阐述了人工智能算法黑箱的法律规制问题；陈思[24]提出技术异化的风险，并从技术、价值两个层面探讨合理有效的人工智能算法治理方案。在国外，J. Yang[25]分析了人工智能算法偏差和不透明性对法律决策的影响及其规律；H. W. Liu 等[26]通过对“美国威斯康星州诉卢米斯案”详细剖析“算法化”所带来的风险，并提出了改进人工智能决策责任的方法；I. Giuffrida[27]针对人工智能算法带来的法律与伦理责任进行研究；A. Simoncini[28]描绘了人工智能算法与法律之间的内在张力，对民主政策为基本自由提供有效保护而制定的标准进行了审查和批评，并提出“预防性宪政”理论的建议；F. J. Z. Borgesius[29]评估了欧洲目前针对歧视性算法判决的法律保护，并对改进现行文件的执行规则提出建议。综上所述，图情领域学者对算法所带来安全风险的关注较少，尤其缺少突出情报领域特色的成果。因此本文基于对智能情报分析核心算法及研究问题的阐释，结合实际应用场景对算法所带来的安全风险及算法风险各因素间交互关系进行分析，最后提出了算法风险前瞻预防与规制的对策建议。该研究不仅能够为政府防范与化解算法所带来的安全风险与误判提供决策依据，还能使国家情报工作战略需求与战略导向适应总体国家安全观。

2　智能情报分析核心算法与研究问题

情报分析是对多源数据进行汇总、处理、评估和分析后将结果转化为情报的过程，传统情报分析方法主要以人工分析和检索式分析为主[30]。随着新兴技术的快速发展，面对海量数据人工智能凭借自我学习的强大优势，极大提升了情报分析的能力，使得情报分析逐步向知识驱动和数据驱动融合发展方向转变[31]。正是基于此，智能情报分析在反恐、金融、军事、安全、竞争、应急等情报工作领域得到了广泛的应用，所用到的核心算法和主要研究问题具有共性特点，而对这些共性特点的研究有助于认识智能情报分析中的算法风险。

2.1　智能情报分析核心算法

智能是人类智能与人工智能的混合产物[32]，智能情报分析是以情报领域专家知识与经验为基础利用算法和数据所形成的智能计算程序。智能情报分析中算法具有较强的自我学习、自我纠错和自我完善能力，本文将其划分为五类[33]：机器学习、深度学习、知识图谱、自然语言处理和计算机视觉，见图 1。

（1）机器学习是智能情报分析应用场景中最基础的算法，它建立在统计学理论基础之上，让机器模拟人类来进行自我学习，总结规律与特征，不断地优化迭代挖掘海量数据中所蕴含的情报。主要包括分类算法、回归算法、聚类算法、降维算法、概率图模型算法、优化算法等。

（2）深度学习算法是对机器学习算法进阶式研究，通过建立类似于人脑的分类模型结构，对输入数据逐级提取从底层到高层的特征，进而建立语义映射关系。它包括的反向传播神经网络、卷积神经网络、深度神经网络、循环神经网络和长短记忆网络等是情报挖掘与预测的核心算法，它常与计算机视觉、自然语言处理相结合，此类算法具有复杂度高、透明度低等特点。

（3）知识图谱算法在聚合信息、知识表达、自动推

理方面有着独特的优势,它能从数据中识别、发现、推断事物之间的复杂关系并构建情报中的知识体系。它包括的抽取算法、路径查找算法、中心度算法、图算法等是智能情报关联、情报可视化应用中核心算法。

(4)自然语言处理算法将人类语言转化为计算机可识别的语言,并能够实现人机之间用自然语言进行通信的技术,它是让计算机形成能够模拟人脑结构的人工神经网络,通过加工处理符号信息来实现语义的理解转换。它包括语法分析、句法分析、文本向量化、语义分析等。其中语义分析是很多领域智能情报分析应用场景的核心算法,其精准度对实际应用有着重要的影响。

(5)计算机视觉算法是情报感知与表达的过程,让机器拥有对海量图像数据提取、理解和分析的能力,它在对图像和视频情报的识别、跟踪、测量及处理的过程体现了巨大的优势。它包括的图像分类算法、对象监测算法、目标跟踪算法、语义分割算法、实例分割算法等在基于图像、视频大数据的智能情报分析中应用较为广泛。

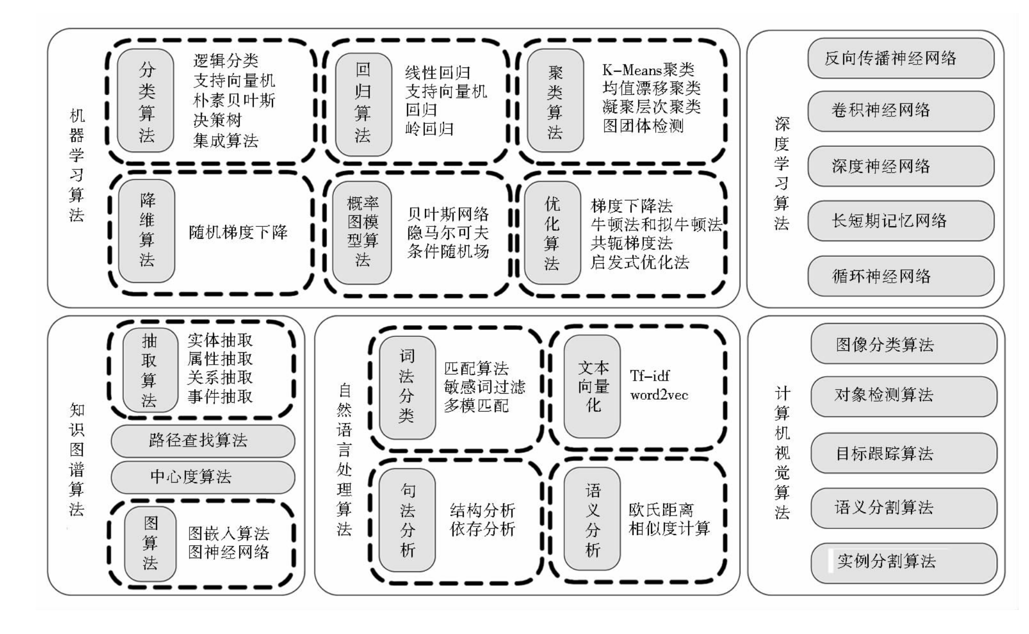


图1　智能情报分析核心算法分类

算法间使用有着繁复的交互关系,往往在智能情报分析实际应用中它们经常混合使用,这使得算法在更聪明的同时也会变得难以理解,尤其是自然语言处理中语义分析算法所形成的意图识别、模糊关联、推理判断等功能,以及深度学习中神经网络算法所具有的多神经元、分布式并行计算、多层深度反馈调整等特点,这加剧了算法的复杂性和不透明性。

2.2　智能情报分析研究问题

智能情报分析能够简化数据处理流程,让情报分析人员有更多的时间和精力学习并运用专业知识,为用户提供高水平、快速、有价值的分析结果[34]。虽然人工智能无法简化情报分析所有复杂性,但可以智能化完善分析的关键环节,为分析判断提供依据[35]。在赵志耘等提出的基于模型的情报分析[36]和化柏林等提出的智能情报分析系统功能结构[10]基础上,本文认为对智能情报分析的研究主要包括以下六个方面:智能情报感知、智能数据采集、智能情报推荐、智能情报关联、智能情报预测和智能情报解读,如图2所示:

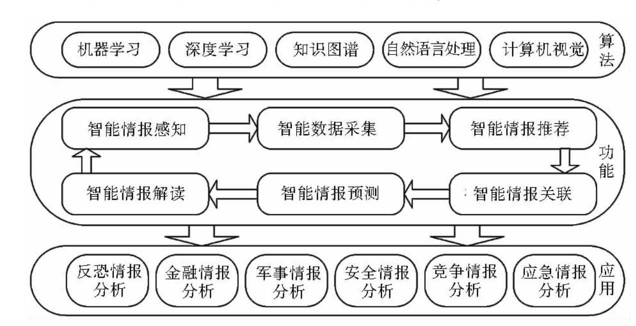


图2　智能情报分析主要研究问题架构

(1)智能情报感知是针对应用场景的变化实时感知用户需求。它通过机器学习算法优化数据流量,短时间内过滤无效的数据,实现从数据构建向数据智能感知转换,对潜在需求进行早期预警与监测。

(2)智能数据采集是利用深度学习、知识图谱、自然语言处理算法从多渠道进行数据实时采集,在知识表示的基础上为用户提供支持语义分析与意图理解的探索

式采集模式,进而实现智能识别数据源特征、自适应采集规则、动态调整采集策略、数据筛选与甄别等功能。

(3)智能情报推荐是情报获取的重要手段之一,在智能情报分析过程中除了智能采集的信息外,还需要在专业的情报库中识别出关联性强的情报。因此智能推荐就会发挥重要作用,它通过智能推荐深度学习、自然语言处理等算法智能感知情报需求并向情报人员推送相关情报,知识图谱引入了更多的语义关系,以实现智能情报推荐的精准性。

(4)智能情报关联是能够通过强化知识的关联性来发现和揭示数据中隐含的情报[37]。基于知识图谱和自然语言处理中的一系列算法将不同数据组合起来,支持对单一实体的语义检索与动态扩展机制,形成映射关联,进而实现对单点实体知识网络的递进式与个性化的可视化情报分析。

(5)智能情报预测是对情报进行智能化的预测分析。通过情报数据对比和历史情报数据测量,基于知识图谱、自然语言处理、神经网络等算法实现情报演化分析、情报发展路径分析和情报发展趋势预测分析。

(6)智能情报解读是对情报分析结果智能化的解读判断。它利用深度学习算法发现数据的规律,探寻相关因素分析数据现象及现象背后的原因,在不断地累积情报数据、行业规则、分析模式的过程中形成智能化的情报解读方案,利用自然语言处理中的语义分析算法辅助情报分析人员生成情报报告,不同的情报报告影响着情报需求,使得情报需求反作用于智能情报感知,进而形成了智能情报分析研究的循环性。

随着数据呈指数级增长,加之深度学习中多层神经网络算法结构复杂,使得近些年在智能情报分析实际应用案例中不断爆出算法背后隐藏的安全风险,这些安全风险引发了社会民众的担心和质疑。

3 智能情报分析中算法风险研究

以算法为基础的智能情报分析既是促进国家情报工作发展的强大动力,也是人工智能时代情报竞争中博弈的武器,但智能情报分析不能一直停留到国家安全形式逻辑框架下,要实现主观和客观相结合的全流程分析,事实上由算法导致智能情报分析的局限性、主观性、安全性等问题较为突出,这与国家安全的总体性原则不相符,长此以往由算法所引发的各类风险随之而来。知识、数据、算法、算力是人工智能不可或缺的要素,其中算法的安全性和鲁棒性是人工智能核心问题。要想真正发挥智能情报分析在情报工作的重要作用,就必须要认识算法风险、分析算法风险成因及其严重后果。文章基于近年来智能情报分析实际应用场景对算法风险进行剖析。

3.1 算法黑箱,使情报分析不透明

3.1.1 认识算法黑箱

算法黑箱是指算法能被看到、理解的只有输入和输出两个环节,内部处理过程完全不透明、不公开,使用户无法获知算法最终目标和意图。算法黑箱不只意味着不能观察,还意味着即使计算机试图向用户解释,用户也无法理解。智能投顾是金融情报分析领域重要应用之一,它运用深度学习、知识图谱等算法及自动化管理技术,在对金融市场进行分析的基础上,通过获取用户的风险偏好水平及大致预期收益率等指标来生成用户画像,再搜选多维数据帮助用户形成主、被动投资策略相结合的定制化投资组合建议[38],在此过程中算法完全是不透明的。然而,很多用户不清楚投资建议是如何形成的,更没有意识到实现投顾智能化和自动化的前提和基础是以技术理性替代人类思维,这意味着“算法”将在很大程度上成为智能投顾所遵循的基本逻辑。宾夕法尼亚州法学院的 Tom Baker 和荷兰鹿特丹伊拉斯谟大学的 Dellaert 教授认为:公众不能预设智能投顾机器人没有人类所具有的不纯动机[39]。由此可见算法缺少透明性或解释性机制使得智能投顾蕴藏着巨大的安全风险。

3.1.2 算法黑箱的成因

基于对智能投顾分析发现算法黑箱形成的原因有二:①算法复杂性导致形成算法黑箱是技术因素的体现。近年来深度学习的快速发展使得算法不遵循传统机器学习的解释过程,而是由计算机基于最原始数据直接自动学习形成高级情报结果,这使智能投顾整个过程缺少透明性或解释性机制,甚至分析人员也无法解释,进而形成了算法黑箱,尤其是深层神经网络的兴起与运用,更加凸显了算法黑箱现象带来的技术屏障,使得算法变得更加难以识别。②利益导致形成算法黑箱是社会因素的体现。在智能投顾中技术公司因负责算法的设计而有意掩饰算法的运行过程,使用户和社会公众几乎完全置身算法黑箱之中。通常来说这都是利益引发的黑箱,算法一旦透明可能会引发技术公司经济利益受损:一是担心智能投顾算法的知识产权被侵犯。算法的研发与运行长期以来属于情报工作领域机密,一旦透明可能会使得其核心理念被盗用,快速被竞争者复制或效仿,进而使得知识产权被侵犯;二是担心智能投顾的分析过程透明导致情报泄露。算法是智

能投顾的核心，算法公开使得智能投顾的分析过程完全透明，可能会导致出现情报泄露的严重后果。

3.1.3 算法黑箱的风险

算法黑箱的确给智能情报分析的良性发展带来了严峻挑战，它使情报分析过程中的行为缺乏有效解释力，一旦算法被封闭为"黑盒子"，对于用户和情报分析工作来说将是令人极度恐慌的，分析结果会出现不确定性、不透明性、高风险性，在智能投顾中算法黑箱带来的风险包括以下两点：①用户画像主要依靠算法对投资者的行为数据、社交情况、交易记录、财务状况等数据进行汇集、筛选和建模分析，在此过程中不但用户的个人隐私在不经意间就受到了侵犯，如果算法基于片面性或歧视性的指标对用户进行画像，这不仅没有真正解决金融资源配置的不平等问题，反而在技术层面还会形成固化的歧视和排斥。②算法在智能投顾中执行过程过于复杂，且缺少透明性或解释性机制，这可能使其直接掩盖金融投资中潜在的风险。随着宏观市场的走向一旦分析人员和用户走入迷雾，形成过度的投资预期，并采取较为激进的风险投资行动，造成金融市场失衡，进而影响社会稳定。这个过程中他们不了解算法执行过程，甚至他们都不清楚什么是安全的，什么是有风险的，更加难以辨别真伪。

3.2 算法缺陷，陷入"情报"茧房

3.2.1 认识算法缺陷

算法歧视带来的不平等、算法偏见所带来的不公正等现象都属于算法缺陷范畴，算法缺陷是指算法本身存在欠缺或不够完备的地方[40]，它是算法不鲁棒的主要表现。美国威斯康星州诉卢米斯案中引入了COMPAS辅助法官断案受到了热议，COMPAS是根据对犯罪者的访谈和来自司法部门的情报，利用深度学习、知识图谱算法来推理预测其再犯的风险，旨在帮助法官做出智能化的司法决策。此案中COMPAS分析显示卢米斯"暴力风险高，再犯风险高，预审风险高"，最终法庭将卢米斯确定为对社区构成高风险的人，判处其六年有期徒刑和五年的延期监督。让社会备受质疑的是COMPAS算法中将种族、性别等带有歧视性因素纳入考量范围，这严重侵害了卢米斯正当程序权利，虽然法院已对卢米斯案进行了释疑，但仍不能平息社会舆论对此问题的争论，卢米斯案及其相关论争反映了算法歧视与程序正义的复杂问题[41]。目前在企业招聘[42]、智能推荐[43]等场景中深度学习、语义分析算法都发挥了重要作用，但在这些应用中也都不同程度的存在由算法引发的不公平、不公正问题。2018年美国皮尤研究中心发布的《公众对计算机算法的态度》报告则显示，约六成美国公众认为算法总有偏见[44]。由此可见算法带来的偏见与歧视问题受到社会关注与民众的质疑。

3.2.2 算法缺陷的成因

基于对上述应用场景分析发现算法缺陷形成的原因有二：①非主观性算法缺陷。由于部分程序人员经验受限形成了非主观性算法缺陷，人工智能算法具有复杂性高的特点，这就对程序人员的专业素养要求较高，部分程序人员一味注重分析结果，而并未意识到所用算法的潜在风险，正是在诉卢米斯案、企业招聘等应用场景中，程序设计人员基于历史数据将种族、性别等因素纳入考量范围，但他们并未意识到结果会带有歧视与偏见，致使形成了非主观性算法缺陷。②主观性算法缺陷。在政治利益、经济利益等因素驱动下导致出现了主观性的算法缺陷。开发者或设计者可能会将自己的歧视或偏见带入算法中，而算法也将这种歧视或偏见延续下去。如在智能推荐场景中，推荐算法的本质就是用过去的经验做预测并推送，而过去的经验中原本就带有出于某种目的歧视或偏见可能会在算法中固化并在未来得到强化和扩大，致使形成了主观性算法缺陷。

3.2.3 算法缺陷的风险

算法本身的复杂性特征和算法黑箱造成了透明度缺失，共同导致了用户根本无法获知算法内部是否存在缺陷，算法缺陷会使分析结果的安全性、准确性、全面性出现偏差，它可能会引发整个社会出现歧视或偏见的重大安全风险。在诉卢米斯案、企业招聘、智能推荐应用场景中的算法均已经体现出分析者对性别、种族等因素的先入之见，这种算法缺陷会使决策者陷入"情报"茧房当中，强化或过度依赖算法所带来的情报分析便捷可能催化极端倾向，其安全风险尤甚。本文对"情报"茧房现象进行如下解释，如图3所示，基于马斯洛需求理论[45]在智能推荐算法不断作用下，使情报人员获取信息从多源到单一，如果情报人员以正反馈的形式依赖存在缺陷的单一信息，这就容易使其逐渐陷入信息缺失的"茧房"之中，而情报在信息链中是信息的高级阶段，情报是对信息的再加工与处理，基于对信息茧房的认知[46]，情报人员关注的情报领域会习惯性地被自己的主观意识所引导，从而其思想桎梏于像蚕茧一般的"茧房"中。"情报"茧房会使情报人员先入之见逐渐根深蒂固，最终导致出现情报分析结果失准、偏差的风险，管理决策难以形成共识[47]。

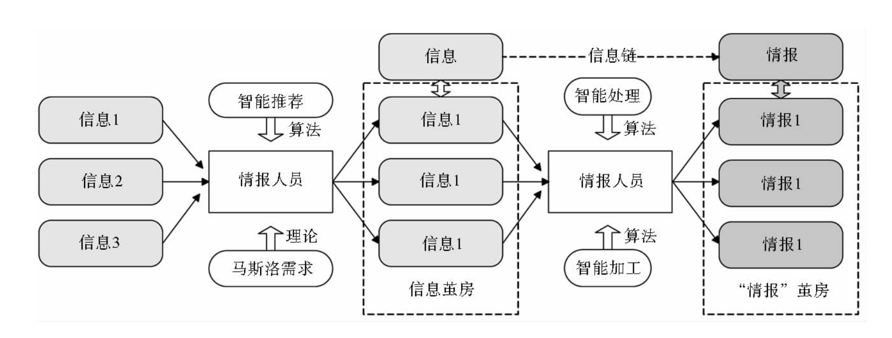


图 3 "情报"茧房形成机理图

3.3 算法操控,使决策行为失控

3.3.1 认识算法操控

算法操控是指算法被人类操控,或算法异化后人类被算法操控,用户要认识并防范各领域智能情报分析中任何形式的算法操控行为。2016 – 2018 年间剑桥分析公司在美国大选、英国脱欧等事件中利用精准营销影响选民政治态度使其名声大噪[48]。剑桥分析公司基于用户行为数据,利用深度学习算法建立模型,分析不同用户群体的人格特质、潜在需求、性格和负面情感等特征,并根据社交媒体上的评价倾向快速识别用户个人隐私信息,从而建立用户档案,成为其对用户行为做出更为准确评估与预测的重要依据。在此过程中选民用户并未认识到算法所带来的潜在风险,但事实上剑桥分析公司依据选民性格和认知特点,利用对算法的操控向用户精准投放政治营销广告、制造水军账户发文传播相关政治理念,直接影响到了选民政治态度与投票结果,这种对算法操控的行为不但会加剧新的社会不平等风险,还会危及到国家政治安全。

3.3.2 算法操控的成因

基于对上述应用场景分析及预测发现算法操控形成的原因有二:①算法被人类操控。被利益集团操纵所形成的算法操控风险是主观因素,因为算法主要依赖程序设计人员的判断与选择。T. J. Dunning 在《工联和罢工》一书中提到"资本有 20% 的利润,它就活跃起来;有 50% 的利润,它就铤而走险;为了 100% 的利润,它就敢践踏一切人间法律;有 300% 的利润,它就敢犯任何罪行,甚至冒绞首的危险"[49]。剑桥分析公司也正是出于利益原因对算法进行操控直接影响了选民的投票结果,丑闻被曝光后引发了社会一片哗然,此后剑桥分析公司遭到声讨并破产。②人类被算法操控。"情报"茧房加之算法异化是人类被算法操控的重要成因。算法异化是指人类所创造出的算法,成为一种异己的、敌对的力量,危害社会、反制人类的力量。算法可以通过大数据实现自我学习、自我训练、自我产生的过程,根据人类的个性特点定制符合特定受众观点和喜好的信息进行投放,进而引导并操控人类改变行为,尽管人类仍然参与其中,但算法已然摆脱了需要依赖人类表达能力的局限,从而极大地提升了算法能力并扩展了其应用范围。

3.3.3 算法操控的风险

算法操控由人类对算法操控所引发的算法对人类操控所产生的安全风险更加可怕。①人类操控算法的风险:著名未来学家托夫勒在《权力的转移》中提出人工智能时代,全球性大公司对数据的垄断,将削弱国家的政治聚合力和国家中心意识,可能会导致少数技术超人操控全球经济和政治[50]。剑桥分析公司对算法操控这种行为带来了严重的安全风险,一是会加剧新的社会不平等风险出现,即民众在数字信息领域的"贫富差距"不断拉大,进而演变为政治上的新霸权 – 算法权力;二是民众对政府失去信任,民众成为算法运行中的一个元素,消解民众的主体性地位,忽略其价值,逐渐使得民众对政府及社会失去信任;三是会严重危及到国家政治安全,算法操控直接影响到政治正义,而政治正义是一个国家必须保障的政治基础,一旦该平衡被打破,将会危及到国家政治安全。②算法操控人类的风险:算法异化原则上属于算法缺陷的一种,只不过算法异化后可能会给社会带来更严重的后果。当前算法的飞速发展和自我进化已初步验证了埃鲁尔在其《技术社会》中"技术的发展通常会脱离人类的控制"的预言[51]。而未来算法异化加之"情报"茧房桎梏将操控情报分析人员和决策者的思维,进而使得决策行为被操控,更可怕的是被算法操控的行为人类无法识别,这种不可预知的后果不但危及到社会稳定和国家安全,甚至会打开危害全世界和平的"潘多拉魔盒",这与总体国家安全观战略严重背离。

3.4 算法风险间交互关系

通过对智能情报分析实际应用场景的研究不难发现,算法黑箱、算法缺陷、算法操控之间即具有复杂的

交互关系，又相互作用影响，如图4所示。算法黑箱是算法风险的产生根源，算法复杂性、利益是算法黑箱形成的主要诱因，算法黑箱为算法权力的隐形运作提供了空间和条件，算法缺陷与算法操控也是由算法黑箱所衍生出的风险，原则上算法操控属于算法缺陷，利用算法操控他人活动同样会导致出现算法权力，算法权力是以利益为导向，并具有与生俱来的歧视性和偏见性。政治、经济等利益因素会导致算法被操控，算法被操控务必会使算法存在缺陷，而算法缺陷容易使情报人员陷入茧房之中，一旦出现"情报"茧房加之算法异化就可能会出现人类被算法操控的安全风险，如图4路径线条加粗所示，由算法黑箱－算法缺陷－算法异化（"情报"茧房）－被算法操控，可以清晰地理出这条风险发生线路，这也是算法风险中最令人担忧的。

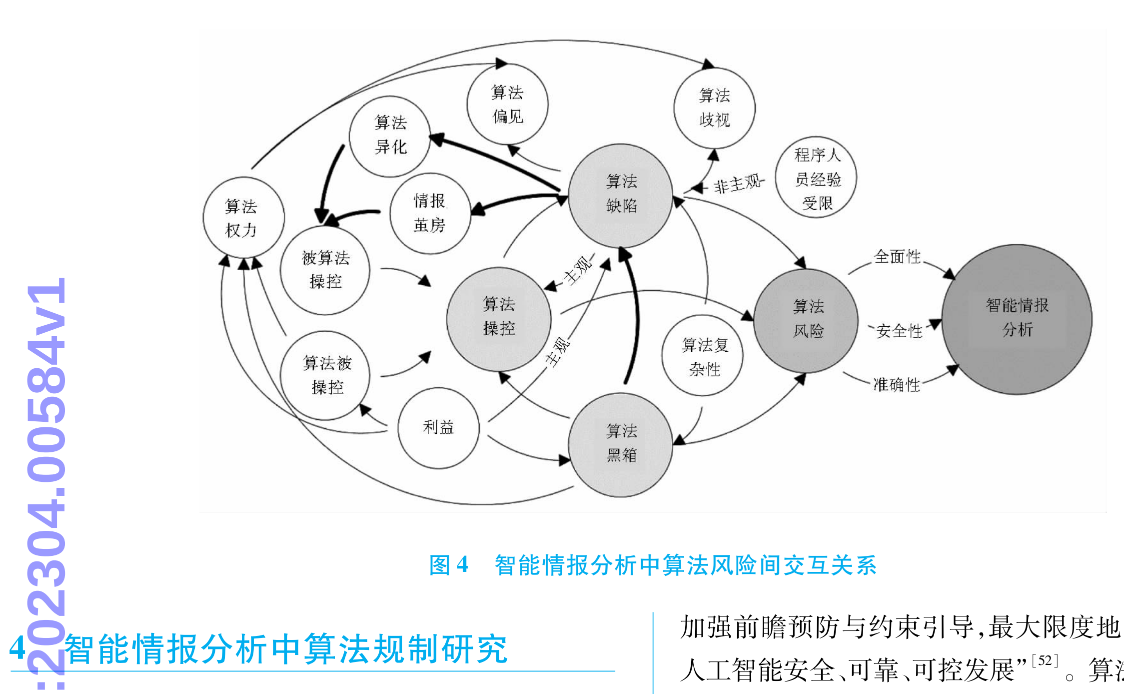


图4　智能情报分析中算法风险间交互关系

4　智能情报分析中算法规制研究

除了以上智能情报分析的应用场景，还有很多情报工作领域一旦出现安全事件其后果不堪设想，如：国家安全情报、军事情报、反恐情报等。因此要特别重视智能情报分析中算法风险，尤其是在对国家安全与发展影响较为严重的情报工作领域更要防范与化解算法风险。在《新一代人工智能发展规划》中提出"在大力发展人工智能的同时，必须高度重视可能带来的安全风险挑战，加强前瞻预防与约束引导，最大限度地降低风险，确保人工智能安全、可靠、可控发展"[52]。算法风险是技术问题，更是社会问题，算法的良性发展不但取决于自身的创新，还取决于社会的正确约束引导。算法规制是防范与化解智能情报分析中算法风险的重要手段，针对算法存在的风险要通过事前评估、事中监管、事后问责的手段建立循序渐进的算法规制框架，从总体安全视角实现算法规制的良性循环与协调发展，以期为政府机构、情报机构制定政策提供对策建议，如图5所示：

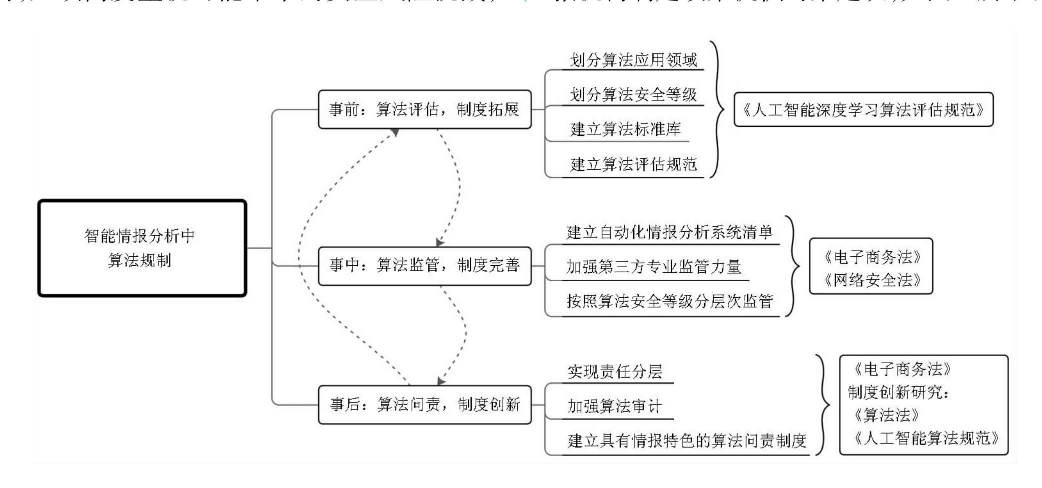


图5　智能情报分析中算法规制框架

4.1 事前:算法评估,制度拓展

算法评估是对算法安全性、鲁棒性、可移植性、效率等系统性的评价,对算法的评估是预测与识别算法风险的有效手段。2018 年 7 月中国电子技术标准化研究院联合多家产学研单位共同编制的《人工智能深度学习算法评估规范》(以下简称《规范》)提出了人工智能深度学习算法的评估指标体系、评估流程及评估内容,其中评估内容包括对需求阶段、设计阶段、实现阶段和运行阶段的评估,《规范》是我国智能算法评估的先导性制度之一,它对其他领域制定算法政策起到重要指导作用。当前我国情报工作领域尚缺少算法评估制度,未来国家情报机构应基于情报工作特点,参照《规范》对智能情报分析进行算法评估制度方面的拓展:①是划分算法应用领域,金融、军事、安全、反恐、应急等情报领域的算法评估内容和侧重点应有所差别,不同领域应有符合自身需求的算法评估机制;②是划分算法安全等级,根据算法复杂性及算法应用后对社会可能产生的负面影响将算法安全等级进行细分;③是建立算法标准库,建立具有"应用领域 – 安全级别 – 最优算法"对应关系的数据库,去除不必要的繁复,以更易理解的方法、算法或模型来代替那些过于复杂且难以解释的方法。④是建立算评估规范,促进情报工作机构及相关组织在不同领域、不同阶段建立有益于算法评估的规范。

4.2 事中:算法监管,制度完善

对算法缺乏有效监管是当前社会所面临的一个棘手难题。算法监管应包括监测、预警、保障等环节,最重要的是通过政策法规手段对算法透明度予以保障,避免对研发者信息披露义务过于宽泛的豁免[53]。2018 年起我国网信管理部门开始重视算法监管问题,起初对算法监管实际是强化现行法律的适用,更好地保护用户权益和网络安全,主要对算法分发规则和伦理道德进行监管,但随着深度学习算法自我学习能力的不断提升,应进一步加强对算法监管的力度。到目前为止虽然我国尚未颁布关于算法监管专门法规,但在《电子商务法》和《网络安全法》中都规定了与算法相关的条款,国家层面已初步将算法规则纳入法律监管体系。基于我国现有较为零散的算法监管政策法规,建议从以下三方面完善情报工作领域算法监管相关制度:①是建立自动化情报分析系统清单,国家情报工作行政管理部门将清单内系统纳入事前评估影响、事中持续监测审查的范围,授予监管机构调查权限以访问相关信息;②是加强第三方专业监管力量,要支持了解情报工作的学术性组织和非营利机构介入,可以通过在情报领域成立智能算法研究委员会等方式扩充专业监管机构,同时协调各方利益加强行业自律;③是按照算法安全等级分层次监管,鉴于情报分析的算法多数都有一定保密级,监管机构或行业组织应要求义务人向监管机构或行业组织备案其算法或算法逻辑,尤其要对产生较大负面影响的算法进行严格监管。

4.3 事后:算法问责,制度创新

算法问责是指情报机构或个人在应用算法过程中对于国家和社会产生的不良影响所应承担的责任以及采取相应救济措施的过程[54]。通过强化算法结果的责任性有助于提升算法分析的公开、公平和向善的价值导向,并不会因为是算法导致分析失误就能规避问责,更要加大算法操控现象的问责力度,算法问责也反向作用于算法评估制度拓展思路。2019 年欧盟出台的《算法责任与透明治理框架》将算法透明度和可问责性作为治理重点,2019 年美国国会引入《算法责任法案》,把透明性作为问责的重要因素,2019 年德国数据伦理委员会发布针对数据和算法的建议,建议将重点放到算法监管和算法责任上来,算法设计要以人为本,与社会核心价值观相符合,注重可持续性、稳定性和安全性。我国在 2018 年制定的《电子商务法》规定了网络平台算法个性化推荐结果的消费者保护义务,对于违反规定的企业,由相关部门责令其限期改正,没收违法所得,并给与相应处罚。截至目前我国尚缺乏系统的算法问责制度,尤其在智能情报分析中,应适当借鉴欧美国家算法问责措施,并从以下方面加强对算法问责制度创新研究:①是实现责任分层,发挥情报机构中学会及委员会的作用,对算法评估和算法监管过程中重点环节制定专项风险点,对于不同算法风险明确责任标准;②是加强算法审计,算法评估和算法审计有着密切的联系,提升审计效率和质量是算法问责的重要关注点;③是建立具有情报特色的算法问责制度,如:国家适时考虑制定《算法法》,国家情报机构针对已有法规制定适用于情报领域的《人工智能算法审查规范》、《算法责任框架》等,在强化问责法律效应的同时,对各领域情报工作起到指导作用。

5 结语

虽然算法对智能情报分析的挑战显而易见,但简单地将算法风险问题归咎于人工智能是不客观的,因为算法本身是无害的,去除政治、经济利益等因素,关键在于如何合理的利用智能算法辅助情报分析。在智能情报分析中算法价值与风险是一个复杂的问题,要在两者间找到协同点以实现共同发展,可以肯定的是

人工智能永远都是辅助人类做出分析,无法替代人类决策,这就需要情报人员具备较高的数据素养和情报素养,以便在情报分析中能够有效的利用智能分析工具和方法,并能向用户解释情报分析过程和最终结果。当下人类正致力于创造人工智能大脑,推动机器人多维度学习,尤其是提高机器人的情感水平,当机器主体具有认知能力、自主性及人类情感时,人类对算法操纵的态度可能发生转变,尊重客观事实或将变成主流趋势。因此,正视智能算法的两面性也是防范与化解算法风险的有效途径。由于情报分析自身具有较为突出的特性,有别于其它领域的信息分析与数据分析,本文对智能情报分析中算法风险及其规制进行初探,以期做出凸显情报特色、突出情报领域话语权的研究。随着智能算法在情报工作的广泛应用,算法风险问题势必会越来越突显,未来研究团队还将继续通过技术和制度手段对算法风险识别、算法治理等问题进行深入研究。

参考文献:

[1] 王忠军,于伟,杨晴. 科技情报机构实践创新发展专家访谈[J]. 情报理论与实践,2017,40(12):145.

[2] 涂元季. 钱学森书信:1993 年 8 月 8 日致戴汝为[M]. 北京:国防工业出版社,2007:311－313.

[3] 王飞跃. 从激光到激活: 钱学森的情报理念与平行情报体系[J]. 自动化学报,2015,41 (6):1053－1061.

[4] 李广建,罗立群. 计算型情报分析的进展[J]. 中国图书馆学报,2019,45(4):29－43.

[5] 李广建,江信昱. 论计算型情报分析[J]. 中国图书馆学报,2018,44(2):4－16.

[6] 陈雪飞,李辉,靳晓宏,等. 计算情报初探[J]. 情报理论与实践,2020,43(3):11－16,70.

[7] 胡昌平,吕美娇. 大数据与智能环境下的情报学理论发展[J]. 情报理论与实践,2020,43(10):1－6.

[8] 栗琳,孙敏. 数据智能技术驱动的情报全流程变革及发展[J]. 情报理论与实践,2020,43(10):7－12.

[9] 邱韵霏,李春旺. 智能情报分析模式:数据驱动型与知识驱动型[J]. 情报理论与实践,2020,43(2): 28－34.

[10] 化柏林,李广建. 智能情报分析系统的架构设计与关键技术研究[J]. 图书与情报,2017(6): 74－83.

[11] 曾文,李辉,李荣,等. 数据工程视角下的智能情报分析与应用探索[J]. 情报理论与实践,2018,41(7): 31－34,59.

[12] 孙建军,李阳. 论情报学与情报工作"智慧"发展的几个问题[J]. 信息资源管理学报,2019,9(1):4－8.

[13] 冯秋燕,朱学芳. 人工智能在情报工作中的应用研究[J]. 情报理论与实践,2019,42(11):27－33.

[14] 牛海波,栗琳. 智能时代情报工作展望[J]. 情报理论与实践,2020,43(1):12－17.

[15] 曾庆华,陈成鑫. 基于综合集成方法的反恐情报分析系统构建[J]. 情报杂志,2018,37(4):27－32.

[16] 丁晓蔚,苏新宁. 基于区块链可信大数据人工智能的金融安全情报分析[J]. 情报学报,2019,38(12):1297－1309.

[17] 王天尧,吴素彬. 人工智能在军事情报工作中的应用现状、特点及启示[J]. 飞航导弹,2020(4):46－51.

[18] 黄云芳,王秉. 智能安全情报分析模型的构建[J]. 情报理论与实践,2020,43(11):59－64.

[19] 唐晓波,郑杜,谭明亮. 融合情报方法论与人工智能技术的企业竞争情报系统模型构建[J]. 情报科学,2019,37(7): 118－124,162.

[20] 曾子明,王婧. 社会计算视角下突发事件智能情报服务研究——以上海外滩踩踏事件为例[J]. 情报杂志,2017,36(11):59－64,77.

[21] 贾开. 人工智能与算法治理研究[J]. 中国行政管理,2019(1): 17－22.

[22] 张爱军,李圆. 人工智能时代的算法权力: 逻辑、风险及规制[J]. 河海大学学报(哲学社会科学版),2019. 21(6):18－24.

[23] 徐凤. 人工智能算法黑箱的法律规制－以智能投顾为例展开[J]. 东方法学,2019(6):78－86.

[24] 陈思. 算法治理: 智能社会技术异化的风险及应对[J]. 湖北大学学报(哲学社会科学版),2020,47(1):78－86.

[25] YANG J. Effects of bias and opacity of artificial intelligence algorithms on legal decision making and its discipline[J]. Korean Lawyers Association journal,2017,66(3): 60－105.

[26] LIU H W, LIN C F, CHEN Y J, et al. Loomis: artificial Intelligence, government algorithmization, and accountability[J]. International journal of law and information technology,2019,27(2): 122－141.

[27] GIUFFRIDA I. Liability for AI decision-making: some legal and ethical considerations[J]. Fordham law review,2019,88(2): 439－456.

[28] SIMONCINI A. The unconstitutional algorithm: artificial intelligence and the future of liberties[J]. Biolaw journal-rivista di biodiritto,2019(1):63－89.

[29] BORGESIUS F J Z. Strengthening legal protection against discrimination by algorithms and artificial intelligence[J]. International journal of human rights,2020,24(10): 1572－1593.

[30] 刘旭东,苏马婧,朱广宇. 基于自然语言处理的多源情报分析系统的研究与设计[J]. 信息技术与网络安全,2019,38(5):17－21.

[31] 张钹,朱军,苏航. 迈向第三代人工智能[J]. 中国科学:信息科学,2020,50(9):1281－1302.

[32] HAYLES N K. How we become posthuman: virtual bodies in cybernetics, literature, and informatics[M]. Chicago: University of Chicago Press,1999:3.

[33] 高航,俞学劢,王毛路. 区块链与人工智能:数字经济新时代[M]. 北京:电子工业出版社,2018.

[34] Analytic edge lever aging emerging technologies trans form intelligence analysis[EB/OL]. [2020－10－26]. https://www.csis.org/analysis/analytic-edge-leveraging-emerging-technologies-transform-intelligence-analysis.

[35] CLARK R M. Intelligence analysis: a target-centric approach[M].

Washington: CQ Press, 2004.

[36] 赵志耘,孙星恺,王晓,等. 组织情报组织智能与系统情报系统智能:从基于情景的情报到基于模型的情报[J]. 情报学报,2020,39(12):1283-1294.

[37] 王军平,张文生,王勇飞,等. 面向大数据领域的事理认知图谱构建与推断分析[J]. 中国科学:信息科学,2020,50(7):988-1002.

[38] 蒋辉宇. 论智能投顾技术性风险的制度防范[J]. 暨南学报(哲学社会科学版),2019,41(9): 48-58.

[39] 刘元兴. 智能金融的"算法可解释性"问题[J]. 金融科技观察,2018(13):1.

[40] 李雨珂. 基于缺陷分析的信息系统质量改进研究[D]. 大连:大连海事大学,2019.

[41] 江溯. 自动化决策、刑事司法与算法规制——由卢米斯案引发的思考[J]. 东方法学,2020(3):76-88.

[42] Why it's totally unsurprising that amazon's recruitment AI was biased against women, business insider[EB/OL]. [2021-05-09]. https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandrawachter-2018-10.

[43] Discrimination in online ad delivery, social science research network[EB/OL]. [2021-05-09]. https://papers.ssrn.com/abstract=2208240.

[44] Public attitudes toward computer algorithms[EB/OL]. [2020-10-26]. https://www.larrysworld.com/public-attitudes-toward-computer-algorithms/.

[45] 胡万钟. 从马斯洛的需求理论谈人的价值和自我价值[J]. 南京社会科学,2000(6):25-29.

[46] 王益成,王萍,王美月,等. 信息运动视角下内容智能分发平台突破"信息茧房"策略研究[J]. 情报理论与实践,2018,41(5):114-119.

[47] 彭兰. 假象、算法囚徒与权利让渡:数据与算法时代的新风险[J]. 西北师大学报(社会科学版),2018,55(5): 20-29.

[48] 穆琳. "剑桥分析"事件"算法黑箱"问题浅析[J]. 中国信息安全,2018(4):92-94.

[49] DUNNING T J. Trades' unions and strikes: their philosophy and intention[M]. London: Knowsley Pamphlet Collection,1860.

[50] 托夫勒. 权力的转移[M]. 刘红,等译. 北京:中共中央党校出版社. 1991.

[51] ELLUL J. The Technological Society [M]. New York: Vintage Books,1964.

[52] 张涛,马海群. 基于文本相似度计算的我国人工智能政策比较研究[J]. 情报杂志,2021,40(1):39-47,24.

[53] LUCERO K. American algorithmic governance policy and implementation approach[J]. Global law review,2020,42(3):5-26.

[54] 迪莉娅. 大数据算法决策的问责与对策研究[J]. 现代情报,2020,40(6):122-128.

作者贡献说明:

张涛:负责论文思路及框架构建,主体内容撰写;

马海群:负责论文内容完善,撰写过程中提出修改意见。

Research on Algorithm Risk and Regulation in Intelligent Intelligence Analysis

Zhang Tao[1] Ma Haiqun[2]

[1] School of Information Management, Heilongjiang University, Harbin 150080

[2] Research Center of Information Resource Management, Heilongjiang University, Harbin 150080

Abstract: [**Purpose/significance**] In recent years, artificial intelligence has brought changes in thinking, concepts, methods and techniques to national intelligence work, making intelligent intelligence analysis gradually become one of the important tasks for the innovation and development of national intelligence. This paper studies the algorithm risk and its regulation in intelligent intelligence analysis to avoid the security risks caused by artificial intelligence algorithms and reduce the factors that restrict the development of intelligence work. [**Method/process**] Based on the interpretation of the core algorithms and research issues of intelligent intelligence analysis, combined with actual application scenarios, the causes and consequences of algorithmic risk and the interaction between the factors of algorithmic risk are analyzed. Establish a gradual regulatory framework for intelligent intelligence analysis algorithms through pre-assessment, mid-event supervision, and post-event accountability. [**Result/conclusion**] This article analyzes the formation mechanism of the "intelligence" cocoon room, and proposes a virtuous cycle of algorithm regulation and coordinated development before, during, and after the risks of algorithm black boxes, algorithm defects, and algorithm manipulation in intelligent intelligence analysis, and believes that Facing the two sides of the algorithm is also an effective way to prevent and defuse the risk of the algorithm.

Keywords: intelligent intelligence analysis algorithmic risk algorithmic regulation intelligent algorithm "intelligence" cocoon room